Guidance for Industry

Average, Population, and Individual Approaches to Establishing Bioequivalence

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register.*

For questions regarding the content of this draft document, contact Mei-Ling Chen, 301-827-5919.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
August 1999
BP

J:\GUIDANC\1716DFT.WPD
8/18/99

Guidance for Industry

Average, Population, and Individual Approaches to Establishing Bioequivalence

Additional copies are available from:

Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
5600 Fishers Lane
Rockville MD 20857
(Tel) 301-827-4573

(Internet) http://www.fda.gov/cder/guidance/index/htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
August 1999
BP

J:\IGUIDANC\1716DFT.WPD
8/18/99

Table of Contents

Draft - Not for Implementation

DRAFT GUIDANCE FOR INDUSTRY[1]

Average, Population, and Individual Approaches to Establishing Bioequivalence

I. INTRODUCTION

This guidance provides recommendations to sponsors of investigational new drug applications (INDs), new drug applications (NDAs), and abbreviated new drug applications (ANDAs) who intend, either before or after approval, to perform in vivo or in vitro bioequivalence (BE) studies. The Center for Drug Evaluation and Research (CDER) recommends three criteria for these comparisons: average, population, and individual criteria. Information about when to use these criteria in BE studies is provided in this guidance and will also be provided in other Agency BE guidances. This guidance focuses on how to use each criterion once a specific criterion has been chosen. When finalized, this guidance will replace a prior FDA guidance entitled *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design*, which was issued by CDER in July 1992 (1992 guidance).

II. BACKGROUND

A. General

Requirements for submitting bioavailability (BA) and BE data in NDAs, ANDAs, and supplements, the definitions of *bioavailability* and *bioequivalence*, and the types of in vivo studies that are acceptable to establish BA and BE are set forth in 21 CFR part 320. This guidance is one of a set of seven planned core BA and BE guidances that provide recommendations on how to meet provisions of part 320 for not only orally administered drug products but also drug products for local action. Draft guidances have been made available for

---

[1] This guidance has been prepared by the Population and Individual Bioequivalence Working Group of the Biopharmaceutics Coordinating Committee in the Office of Pharmaceutical Science, Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration (FDA). This guidance document represents the Agency's current thinking on the approaches for evaluating bioequivalence studies. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes, regulations, or both.

public comment or are under preparation on the following topics:

- BA and BE studies for orally administered drug products — general considerations (under preparation)
- Average, population, and individual approaches to establishing bioequivalence (this guidance)
- Topical dermatological drug product NDAs and ANDAs - bioavailability, bioequivalence, in vitro release, and associated studies (draft published July 1998)
- Bioanalytical methods validation for human studies (draft published January 1999)
- Waiver of in vivo bioavailability and bioequivalence studies for immediate release solid oral dosage forms containing certain active moieties/active ingredients based on a biopharmaceutics classification system (draft published January 1999)
- Bioavailability and bioequivalence studies for nasal aerosols and nasal sprays for local actions (draft published June 1999)
- Bioavailability and bioequivalence studies for oral inhalation drug products for local action: MDIs and DPIs (under preparation).

Defined as *relative BA*, BE involves comparison between a test (T) and reference (R) drug product, where T and R can vary, depending on the comparison to be performed (e.g., to-be-marketed dosage form versus clinical trial material, generic drug versus reference listed drug, drug product changed after approval versus the drug product before the change). Although BA and BE are closely related, BE comparisons rely on a criterion, a predetermined BE limit, and calculation of a confidence interval for the criterion. BE comparisons could be indicated in the presence of certain pharmaceutical product line extensions, such as additional strengths, new dosage forms (e.g., changes from immediate release to extended release), and new routes of administration. In these settings, use of some of the approaches described in this guidance could help determine BE. The approaches discussed in this guidance may also be useful when performing equivalence comparisons in clinical pharmacology and assessing pharmaceutical equivalence.

B. Statistical

In the July 1992 guidance on *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design*, CDER recommended that a standard in vivo BE study design be based on the administration of either single or multiple doses of T and R products to healthy subjects on separate occasions, with random assignment to the two possible sequences of drug product administration. The 1992 guidance further recommended that statistical analysis for pharmacokinetic parameters, such as area under the curve (AUC) and peak concentration (Cmax), be based on a test procedure termed the *two one-sided tests*

procedure to determine whether the average values for pharmacokinetic parameters measured after administration of T and R products were comparable. This approach is termed *average bioequivalence* and involves the calculation of a 90% confidence interval for the ratio of the averages of T and R products. To establish BE, the calculated confidence interval should fall within a BE limit, usually 80-125% for the ratio of the product averages.[2] In addition to this general approach, the 1992 guidance provided specific recommendations for (1) logarithmic transformation of pharmacokinetic data, (2) methods to evaluate sequence effects, and (3) methods to evaluate outlier data.

Based on extensive intramural and extramural discussions, we now recommend that the average BE be supplemented by two new approaches, termed *population* and *individual bioequivalence*. Supplementation with these further criteria is suggested because of the limitations of the currently applied average BE criterion. The average BE approach focuses only on the comparison of population averages of a BE measure of interest and not on the variances of the measure for T and R products. The average BE method does not assess a subject-by-formulation interaction variance, that is, the variation in the average T and R difference among individuals. In contrast, population and individual BE approaches include comparisons of both the averages and variances of the study measure. The population BE approach assesses the total variability of the measure in the population. The individual BE approach assesses within-subject variability for T and R product, as well as the subject-by-formulation interaction.

The population and individual approaches reflect differences in the objectives of BE testing at various stages of drug development. These differences are embodied in the concepts of *prescribability* and *switchability*. *Prescribability* refers to the clinical setting in which a practitioner prescribes a drug product to a patient for the first time. In this setting, the prescriber relies on an understanding that the average performance of the drug product has been well characterized and relates in some definable way to the safety and efficacy information from clinical trials. *Switchability* refers to the setting in which a practitioner transfers a patient from one drug product to another. This situation arises with generic substitution, as well as with certain postapproval changes by an innovator or generic firm in the formulation and/or manufacture of a drug product. Under these circumstances, the prescriber and patient should be assured that the newly administered drug product will yield comparable safety and efficacy to that of the product for which it is being substituted.

---

[2] For a broad range of drugs, a BE limit of 80 to 125% for the ratio of the product averages has been adopted for use of an average BE criterion. Generally, the BE limit of 80 to 125% is based on a clinical judgment that a test product with BA measures outside this range should be denied market access. In specified circumstances, clinical judgment can allow widening or narrowing of the BE limit (e.g., 90-111% for narrow therapeutic range drugs/drug products).

III. STATISTICAL MODEL

Statistical analyses of BE data are typically based on a statistical model for the logarithm of the BA measures (e.g., AUC and Cmax). The model is a mixed-effects or two-stage linear model.[3] Each subject, j, theoretically provides a mean for the log-transformed BA measure for each formulation, μ_{Tj} and μ_{Rj} for the T and R formulations, respectively. The model assumes that these subject-specific means come from a distribution with population means μ_T and μ_R and between-subject variances σ_{BT}^2 and σ_{BR}^2, respectively. The model allows for a correlation, ρ, between μ_{Tj} and μ_{Rj}. The subject-by-formulation interaction variance component (Schall and Luus 1993), σ_D^2, is related to these parameters as follows:

$$\sigma^2 = \text{variance of } (\mu_{Tj}-\mu_{Rj})$$

$$= (\sigma_{BT} - \sigma_{BR})^2 + 2\,(1-\rho)\sigma_{BT}\sigma_{BR} \qquad \textbf{Equation 1}$$

For a given subject, the observed data for the log-transformed BA measure are assumed to be independent observations from distributions with means μ_{Tj} and μ_{Rj} and within-subject variances σ_{WT}^2 and σ_{WR}^2. The total variances for each formulation are defined as the sum of the within- and between-subject components (i.e., $\sigma_{TT}^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ and $\sigma_{TR}^2 = \sigma_{WR}^2 + \sigma_{BR}^2$). For analysis of crossover studies, the means are given additional structure by the inclusion of period and sequence effect terms.

IV. BIOEQUIVALENCE CRITERIA

The general structure of a BE criterion is that a function (Θ) of population parameters should be demonstrated to be no greater than a specified value (θ). Using the terminology of statistical hypothesis testing, this is accomplished by testing the hypothesis H_0: $\Theta>\theta$ versus H_A: $\Theta\leq\theta$ at a desired level of significance, often 5%. Rejection of the null hypothesis H_0, that is, demonstrating that the estimate of Θ is statistically significantly less than θ, results in a conclusion of BE. The choice of Θ and θ differs in average, population, and individual BE criteria.

A general objective in assessing BE is to compare the log-transformed BA measure after administration

---

[3] If $Y_{\tau jk}$ denotes the kth observation on the jth individual and τ is the formulation, T or R, then

$$Y_{\tau jk} = \mu_{\tau j} + \epsilon_{\tau jk}$$

where $E(\epsilon_{\tau jk}) = 0$, $Cov(\epsilon_{\tau j k}, \epsilon_{\tau' j' k'}) = \sigma_{W\tau}^2\delta_{\tau\tau'}\delta_{jj'}\delta_{kk'}$, $E(\mu_{\tau j}) = \mu_\tau$, $Cov(\mu_{\tau j}, \mu_{\tau' j'}) = \sigma_{B\tau\tau'}\delta_{jj'}$, $\sigma_{B\tau\tau'} = \sigma_{B\tau}^2$ if $\tau = \tau'$, otherwise $\rho\sigma_{B\tau}\sigma_{B\tau'}$; and $\delta_{jj'}$ is equal to 1 if $j = j'$, otherwise 0.

of T and R products. As detailed in Appendix B, population and individual criteria are based on the comparison of an expected squared distance between T and R formulations to the expected distance between two administrations of R formulation. An acceptable T formulation is one where the T-R distance is not substantially greater than the R-R distance. In both population and individual BE criteria, this comparison appears as a comparison to the reference variance, which is referred to as *scaling to the reference variability.*

Population and individual BE criteria, but not the average BE criterion, allow two types of scaling: reference-scaling and constant-scaling. Reference-scaling means that the criterion used is scaled to the variability of the R product, which effectively widens the BE limit for more variable reference products. Although generally sufficient, use of reference-scaling alone could unnecessarily narrow the BE limit for drugs and/or drug products that have a wide therapeutic window and low variability. CDER recommends that the scaling approach be limited for low variability reference products by scaling to a specified constant, rather than to the variability of the reference product. This is called constant-scaling. This guidance therefore recommends a mixed-scaling approach for the population and individual BE criteria (section IV.B and C). With mixed scaling, the reference-scaled form of the criterion is used if the reference product is sufficiently highly variable and otherwise the constant-scaled form is used. An exception is narrow therapeutic window drugs and/or drug products, where reference-scaling is recommended regardless of the reference variability.

A. Average Bioequivalence

The following approach is recommended for average BE:

$$(\mu_T - \mu_R)^2 \leq \theta_A^2 \qquad \textbf{Equation 2}$$

where

μ_T = population average response of the log-transformed measure for T formulation

μ_R = population average response of the log-transformed measure for R formulation

as defined in section III above.

This criterion is equivalent to the more commonly seen form:

$$-\theta_A \leq \mu_T - \mu_R \leq \theta_A \qquad \textbf{Equation 3}$$

and, usually, $\theta_A = \ln(1.25)$.

B. Population Bioequivalence

The following mixed-scaling approach is recommended for population BE (i.e., use the reference-scaled method if the estimate of $\sigma_{TR} > \sigma_{T0}$ and the constant-scaled method if the estimate of $\sigma_{TR} \leq \sigma_{T0}$).

- Reference-Scaled:

$$\frac{(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)}{\sigma_{TR}^2} \leq \theta_p \qquad \textbf{Equation 4}$$

- Constant-Scaled:

$$\frac{(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)}{\sigma_{T0}^2} \leq \theta_p \qquad \textbf{Equation 5}$$

where:

μ_T = population average response of the log-transformed measure for T formulation
μ_R = population average response of the log-transformed measure for R formulation
σ_{TT}^2 = total variance (i.e., sum of within- and between-subject variances) of T formulation
σ_{TR}^2 = total variance (i.e., sum of within- and between-subject variances) of R formulation
σ_{T0}^2 = specified constant total variance
θ_p = BE limit

Equations 4 and 5 represent an aggregate approach where a single criterion on the left-hand side of the equation encompasses two major components: (1) difference between T and R

population averages ($\mu_T - \mu_R$), and (2) difference between T and R total variances ($\sigma_{TT}^2 - \sigma_{TR}^2$). This aggregate measure is scaled to the total variance of the R product or to a constant value (σ_{T0}^2, a standard that relates to a limit for the total variance), whichever is greater.

The determination of both σ_{T0} and θ_P relies on the establishment of standards. The generation of these standards is discussed in Appendix A.

C. Individual Bioequivalence

The following mixed-scaling approach is recommended for individual BE (i.e., the reference-scaled method is recommended if the estimate of $\sigma_{WR} > \sigma_{W0}$ and the constant-scaled method is recommended if the estimate of $\sigma_{WR} \leq \sigma_{W0}$).

- Reference-Scaled:

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2)}{\sigma_{WR}^2} \leq \theta_I \qquad \textbf{Equation 6}$$

- Constant-Scaled:

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2)}{\sigma_{W0}^2} \leq \theta_I \qquad \textbf{Equation 7}$$

where:

μ_T	= population average response of the log-transformed measure for T formulation
μ_R	= population average response of the log-transformed measure for R formulation
σ_D^2	= subject-by-formulation interaction variance component
σ_{WT}^2	= within-subject variance of T formulation
σ_{WR}^2	= within-subject variance of R formulation
σ_{W0}^2	= specified constant within-subject variance
θ_I	= BE limit

Equations 6 and 7 represent an aggregate approach where a single criterion on the left-hand

side of the equation encompasses three major components: (1) difference between T and R population averages ($\mu_T - \mu_R$), (2) subject-by-formulation interaction (σ_D^2), and (3) difference between T and R within-subject variances ($\sigma_{WT}^2 - \sigma_{WR}^2$). This aggregate measure is scaled to the within-subject variance of the R product or to a constant value (σ_{W0}^2, a standard that relates to a limit for the within-subject variance), whichever is greater.

The determination of both σ_{W0} and θ_I relies on the establishment of standards. The generation of these standards is discussed in Appendix A.

V. STUDY DESIGN

A. Experimental Design

1. Replicated Crossover Designs

Replicated crossover designs can be used irrespective of which criterion is selected to establish BE, although they are not necessary when an average or population criterion is used. Replicated crossover designs are recommended when an individual BE criterion is used to allow estimation of within-subject variances for T and R measures and the subject-by-formulation interaction variance component. The following four-period, two-sequence, two-formulation design is recommended for replicated BE studies (see Appendix B for further discussion of replicated crossover designs):

		Period			
		1	2	3	4
Sequence	1	**T**	**R**	**T**	**R**
	2	**R**	**T**	**R**	**T**

For this design, the same lots of T and R formulations should be used for the replicated administration. Each period should be separated by an adequate washout period.

Other replicated crossover designs are possible. For example, a three-period design, as shown below, could be used:

		Period		
		1	2	3
Sequence	1	**T**	**R**	**T**
	2	**R**	**T**	**R**

A greater number of subjects would be encouraged for the three-period design compared to the recommended four-period design to achieve the same statistical power to conclude BE (see Appendix C).

2. Nonreplicated Designs

A conventional nonreplicated design, such as the standard two-formulation, two-period, two-sequence crossover design, can be used to generate data for BE comparisons where an average or population BE criterion is used. Under certain circumstances, parallel designs can also be used.

B. Study Population

Unless otherwise indicated by a specific guidance, subjects recruited for in vivo BE studies should be 18 years or older and capable of giving informed consent. An attempt should be made to enter as heterogeneous a study population as possible, with a reasonable balance of males and females, young and elderly, and subjects of differing racial groups. Restrictions to entry into the study should be based solely on safety considerations. In some BE studies, it may be useful to recruit patients for whom the drug product is intended. In this situation, sponsors and/or applicants should attempt to enter patients whose disease process is stable for the duration of the BE study. An IND may be required for some ANDA BE studies to help ensure patient safety (see 21 CFR 320.31). For nonpharmacokinetic measures used to establish BE, information is provided in other guidances.

C. Sample Size and Dropouts

A minimum number of 12 subjects should be included in any BE study. When an average BE criterion is selected using either nonreplicated or replicated designs, methods appropriate to the study design should be used to estimate sample sizes. The number of subjects for BE studies

based on either the population or individual BE approach should be estimated by simulation, because analytical approaches for estimation are not available. Further information on sample size is provided in Appendix C.

Sponsors should enter a sufficient number of subjects in the study to allow for dropouts. Because replacement of subjects during the study could complicate the statistical model and analysis, dropouts generally should not be replaced. Sponsors who wish to replace dropouts during the study should indicate this intention in the protocol. The protocol should also state whether samples from replacement subjects, if not used, will be assayed. If the dropout rate is high and sponsors wish to add more subjects, a modification of the statistical analysis may be recommended. Additional subjects should not be included after data analysis unless the trial was designed from the beginning as a sequential or group sequential design.

VI. STATISTICAL ANALYSIS

The following sections provide recommendations on statistical methodology for assessment of average, population, and individual BE.

A. Logarithmic Transformation

1. General Procedures

This guidance recommends that BE measures (e.g., AUC and Cmax) be log-transformed using either common logarithms to the base 10 or natural logarithms (see Appendix D). The choice of common or natural logs should be consistent and should be stated in the study report. The limited sample size in a typical BE study precludes a reliable determination of the distribution of the dataset. Sponsors and/or applicants are not encouraged to test for normality of data distribution after log-transformation, nor should they use normality of data distribution as a reason for carrying out the statistical analysis on the original scale. Justification should be provided if a sponsor and/or applicant believes that their BE study data should be statistically analyzed on the original rather than log scale.

2. Presentation of Data

The drug concentration in biological fluid determined at each sampling time point should be furnished on the original scale for each subject participating in the study. The derived pharmacokinetic measures should also be furnished on the original scale. The

mean, standard deviation, and coefficient of variation for each variable should be computed and tabulated in the final report.

In addition to the arithmetic mean and standard deviation (or coefficient of variation) for T and R products, geometric means (antilog of the means of the logs) should be calculated for selected BE measures. To facilitate BE comparisons, the measures for each individual should be displayed in parallel for the formulations tested. In particular, the ratio of the individual geometric mean of the T product to the individual geometric mean of the R product should be tabulated side by side for each subject. The summary tables should indicate in which sequence each subject received the product.

B. Data Analysis

1. Average Bioequivalence

a. Overview

Parametric (normal-theory) methods are recommended for the analysis of log-transformed BE measures. For average BE using the criterion stated in equations 2 or 3 above, the general approach is to construct a 90% confidence interval for the quantity μ_T-μ_R and to reach a conclusion of average BE if this confidence interval is contained in the interval $[-\theta_A , \theta_A]$. Due to the nature of normal-theory confidence intervals, this is equivalent to carrying out two one-sided tests of hypothesis at the 5% level of significance (Schuirmann 1987).

The 90% confidence interval for the difference in the means of the log-transformed data should be calculated using methods appropriate to the experimental design. The antilogs of the confidence limits obtained constitute the 90% confidence interval for the ratio of the geometric means between T and R product.

b. Replicated Crossover Designs

Linear mixed-effects model procedures, available in PROC MIXED in SAS or equivalent software, should be used for the analysis of replicated crossover studies for average BE. Appendix E includes an example of SAS program statements.

c. Nonreplicated Crossover Designs

For nonreplicated crossover designs, this guidance recommends parametric (normal-theory) procedures to analyze log-transformed BA measures. General linear model procedures available in PROC GLM in SAS or equivalent software are preferred, although linear mixed-effects model procedures may also be indicated for analysis of nonreplicated crossover studies.

For example, for a conventional two-treatment, two-period, two-sequence (2 x 2) randomized crossover design, the statistical model typically includes factors accounting for the following sources of variation: sequence, subjects nested in sequences, period, and treatment. The Estimate statement in SAS PROC GLM, or equivalent statement in other software, should be used to obtain estimates for the adjusted differences between treatment means and the standard error associated with these differences.

d. Parallel Designs

For parallel designs, the confidence interval for the difference of means in the log scale can be computed using the total between-subject variance. As in the analysis for replicated designs (section VI. B.1.b), equal variances should not be assumed.

2. *Population Bioequivalence*

a. Overview

Analysis of BE data using the population criterion (section IV.B) focuses first on estimation of the mean difference between T and R for the log-transformed BA measure and estimation of the total variance for each of the two formulations. This can be done using relatively simple unbiased estimators such as the method of moments (MM). After this is completed, a 95% upper confidence bound for the population BE criterion can be obtained. Population BE should be considered established for a particular log-transformed BA measure if the 95% upper confidence bound is less than or equal to the BE limit, θ_P.

To obtain the 95% upper confidence bound of the criterion, intervals based on validated approaches can be used. Validation approaches should be reviewed with appropriate staff in CDER. Appendix F includes an example of upper confidence bound determination using a population BE approach.

b. Nonreplicated Crossover Designs

For nonreplicated crossover studies, any available method (e.g., SAS PROC GLM or equivalent software) can be used to obtain an unbiased estimate of the mean difference between T and R log-transformed BA measures. The total variance for each formulation should be estimated by the usual sample variance, computed separately in each sequence and then pooled across sequences.

c. Replicated Crossover Designs

For replicated crossover studies, the approach is the same as for nonreplicated crossover designs, but care should be used to obtain proper estimates of the total variances. One approach is to estimate the within- and between-subject components separately, as for individual BE (see section VI.B.3), and then sum them to obtain the total variance. The method for the upper confidence bound should then be consistent with the method used for estimating the variances.

d. Parallel Designs

The estimate of the means and variances from parallel designs should be the same as for nonreplicated crossover designs. The method for the upper confidence bound should be modified to reflect independent rather than paired samples and to allow for unequal variances.

3. Individual Bioequivalence

Analysis of BE data using an individual BE criterion (section IV.C) focuses on estimation of the mean difference between T and R for the log-transformed BA measure, the subject-by-formulation interaction variance, and the within-subject variance for each of the two formulations. For this purpose, we recommend the MM approach (Chinchilli 1996 and Chinchilli and Esinhart 1996). The restricted maximum likelihood (REML) method may be useful to estimate mean differences and variances when subjects with some missing data are included in the statistical analysis.

A key distinction between the REML and MM methods relates to differences in estimating variance terms and is further discussed in Appendix G. After the estimation of the mean difference and the variances has been completed, a 95% upper confidence bound for the individual BE criterion is obtained. Individual BE should be established for a particular log-transformed BA measure if the 95% upper confidence bound is less

than or equal to the BE limit, θ_I.

To obtain the 95% upper confidence bound of the criterion, intervals based on validated approaches can be used. An example is described in Appendix H. Other methods in development may also be useful, and sponsors considering alternative methods are encouraged to discuss their approaches with appropriate CDER review staff prior to submitting their applications or supplements.

VII. MISCELLANEOUS ISSUES

A. Studies in Multiple Groups

If a crossover study is carried out in two or more groups of subjects (e.g., if for logistical reasons, only a limited number of subjects can be studied at one time), the statistical model should be modified to reflect the multigroup nature of the study. In particular, the model should reflect the fact that the periods for the first group are different from the periods for the second group. This applies to all of the approaches (average, population, and individual BE) described in this guidance.

If the study is carried out in two or more groups and those groups are studied at different clinical sites, or at the same site but greatly separated in time (months apart, for example), questions may arise as to whether the results from the several groups should be combined in a single analysis. Such cases should be discussed with appropriate CDER review staff.

A *sequential* design, in which the decision to study a second group of subjects is based on the results from the first group, calls for different statistical methods and is outside the scope of this guidance. Those wishing to use a sequential design should consult appropriate CDER review staff.

B. Carryover Effects

Use of crossover designs for BE studies allows each subject to serve as his or her own control to improve the precision of the comparison. One of the assumptions underlying this principle is that *carryover effects* (also called *residual effects*) are either absent (the response to a formulation administered in a particular period of the design is unaffected by formulations administered in earlier periods) or equal. If carryover effects are present in a crossover study and are not equal, the usual crossover estimate of μ_T-μ_R could be biased. A major limitation of a conventional two-formulation, two-period, two-sequence crossover design is that the only

statistical test available for the presence of unequal carryover effects is the sequence test in the analysis of variance (ANOVA) for the crossover design. This is a between-subject test, which would be expected to have poor discriminating power in a typical BE study. Furthermore, if the possibility of unequal carryover effects cannot be ruled out, no unbiased estimate of μ_T-μ_R based on within-subject comparisons can be obtained.

For replicated crossover studies, a within-subject test for unequal carryover effects can be obtained under certain assumptions. Typically only first-order carryover effects are considered of concern (i.e., the carryover effects, if they occur, only affect the response to the formulation administered in the next period of the design). Under this assumption, consideration of carryover effects could be more complicated for replicated crossover studies than for nonreplicated studies. The carryover effect could depend not only on the formulation that preceded the current period, but also on the formulation that is administered in the current period. This is called a *direct-by-carryover* interaction. The need to consider more than just *simple* first-order carryover effects has been emphasized (Fleiss 1989). With a replicated crossover design, a within-subject estimate of μ_T-μ_R unbiased by general first-order carryover effects can be obtained, but such an estimate could be imprecise, reducing the power of the study to conclude BE.

In most cases, for both replicated and nonreplicated crossover designs, the possibility of unequal carryover effects is considered unlikely in a BE study under the following circumstances:

- It is a single-dose study.
- The drug is not an endogenous entity.
- More than an adequate washout period has been allowed between periods of the study and in the subsequent periods the predose biological matrix samples do not exhibit a detectable drug level in any of the subjects.
- The study meets all scientific criteria (e.g., it is based on an acceptable study protocol and it contains sufficient validated assay methodology).

The possibility of unequal carryover effects can also be discounted for multiple-dose studies and/or studies in patients, provided the studies meet other criteria listed above. Under all other circumstances, the sponsor and/or applicant could be asked to consider the possibility of unequal carryover effects, including a direct-by-carryover interaction. If there is evidence of carryover effects, sponsors should describe their proposed approach in the study protocol,

including statistical tests for the presence of such effects and procedures to be followed.

C. Outlier Considerations

Outlier data in BE studies are defined as subject data for one or more BA measures that are discordant with corresponding data for that subject and/or for the rest of the subjects in a study. Because BE studies are usually carried out as crossover studies, the most important type of subject outlier is the within-subject outlier, where one subject or a few subjects differ notably from the rest of the subjects with respect to a within-subject T-R comparison. The existence of a subject outlier with no protocol violations could indicate one of the following situations.

1. Product Failure

Product failure could occur, for example, when a subject exhibits an unusually high or low response to one or the other of the products because of a problem with the specific dosage unit administered. This could occur with a sustained and/or modified-release dosage form exhibiting dose dumping or a dosage unit with a coating that inhibits dissolution.

2. Subject-by-Formulation Interaction

A subject-by-formulation interaction could occur when an individual is representative of subjects present in the general population in low numbers, for whom the relative BA of the two products is markedly different than for the majority of the population, and for whom the two products are not bioequivalent, even though they might be bioequivalent in the majority of the population.

In the case of product failure, the unusual response could be present for either the T or R product. However, in the case of a subpopulation, even if the unusual response is observed on the R product, there could still be concern for lack of interchangeability of the two products. For these reasons, deletion of outlier values is generally discouraged, particularly for nonreplicated designs. With replicated crossover designs, the *retest* character of these designs can indicate consideration of the deletion of an outlier value. Applicants with these types of datasets should review the outlier possibilities with appropriate review staff.

D. Discontinuity

The mixed-scaling approach has a discontinuity at the changeover point, σ_{W0} (individual BE criterion) or σ_{T0} (population BE criterion), from constant- to reference-scaling. For example, if the estimate of the within-subject standard deviation of the reference is just above the changeover point, the confidence interval will be wider than just below. In this context, the confidence interval could pass the predetermined BE limit if the estimate is just below the boundary and could fail if just above. A solution to this issue is to recommend sponsors use either reference-scaling or constant-scaling at the changeover point. With this approach, the multiple testing inflates the type I error rate slightly, to approximately 6.5%, but only over a small interval of σ_{wR} (about 0.18-0.20). Another solution is to maintain $\sigma_{W0} = 0.2$, but make the switch from constant-scaling to reference-scaling only when the estimate of σ_{WR} is greater than some larger threshold, such as 0.25. Further study of this approach is underway.

REFERENCES

Anderson, S., and W.W. Hauck, 1990, "Consideration of Individual Bioequivalence," *J. Pharmacokin. Biopharm.*, 18:259-73.

Anderson, S., 1993, "Individual Bioequivalence: A Problem of Switchability (with discussion)," *Biopharmaceutical Reports*, 2(2):1-11.

Anderson, S., 1995, "Current Issues of Individual Bioequivalence," *Drug Inf. J.*, 29:961-4.

Chen, M.-L., 1997, "Individual Bioequivalence - A Regulatory Update (with discussion)," *J. Biopharm. Stat.*, 7:5-111.

Chen, M.-L., R. Patnaik, W.W. Hauck, D.J. Schuirmann, T. Hyslop, R.L. Williams, and the FDA Population and Individual Bioequivalence Working Group, "An Individual Bioequivalence Criterion: Regulatory Considerations," *Stat. Med.*, In press.

Chen, M.-L., S.-C. Lee, M.-J. Ng, D.J. Schuirmann, L.J. Lesko, and R.L. Williams, "Pharmacokinetic Analysis of Bioequivalence Trials: Implications for Gender Issues in Clinical Pharmacology and Biopharmaceutics," Submitted.

Chinchilli, V.M., 1996, "The Assessment of Individual and Population Bioequivalence," *J. Pharmacokin. Biopharm.*, 6:1-14.

Chinchilli, V.M., and J.D. Esinhart, 1996, "Design and Analysis of Intra-Subject Variability in Cross-Over Experiments," *Stat. Med.*, 15:1619-34.

Chow, S.-C., 1999, "Individual Bioequivalence - A Review of the FDA Draft Guidance," *Drug Inf. J.*, 33:435-44.

Diletti E., D. Hauschke, and V.W. Steinijans, 1991, "Sample Size Determination for Bioequivalence Assessment By Means of Confidence Intervals," *Int. J. Clin. Pharmacol. Therap.*, 29:1-8.

Efron, B., 1987, "Better Bootstrap Confidence Intervals (with discussion)," *J. Amer. Stat. Assoc.*, 82:171-200.

Efron, B., and R.J. Tibshirani, 1993, *An Introduction to the Bootstrap*, Chapman and Hall, Ch. 14.

Ekbohm, G., and H. Melander, 1989, "The Subject-by-Formulation Interaction as a Criterion for Interchangeability of Drugs," *Biometrics*, 45:1249-54.

Ekbohm, G., and H. Melander, 1990, "On Variation, Bioequivalence and Interchangeability," Report 14, Department of Statistics, Swedish University of Agricultural Sciences.

Endrenyi, L., and M. Schulz, 1993, "Individual Variation and the Acceptance of Average Bioequivalence," *Drug Inf. J.*, 27:195-201.

Endrenyi, L., 1993, "A Procedure for the Assessment of Individual Bioequivalence," in *Bio-International: Bioavailability, Bioequivalence and Pharmacokinetics* (H.H.Blume, K.K. Midha, eds.), Medpharm Publications, 141-6.

Endrenyi, L., 1994, "A Method for the Evaluation of Individual Bioequivalence," *Int. J. Clin. Pharmacol. Therap.*, 32:497-508.

Endrenyi, L., 1995, "A Simple Approach for the Evaluation of Individual Bioequivalence," *Drug Inf. J.*, 29:847-55.

Endrenyi, L., and K.K. Midha, 1998, "Individual Bioequivalence - Has Its Time Come?," *Eur. J. Pharm. Sci.*, 6:271-8.

Endrenyi, L., G.L. Amidon, K.K. Midha, and J.P. Skelly, 1998, "Individual Bioequivalence: Attractive in Principle, Difficult in Practice," *Pharm. Res.*, 15:1321-5.

Endrenyi, L., and Y. Hao, 1998, "Asymmetry of the Mean-Variability Tradeoff Raises Questions About the Model in Investigations of Individual Bioequivalence," *Int. J. Pharmacol. Therap.*, 36:450-7.

Endrenyi, L., Y. Hao, and L. Tothfalusi, "Uncertainty of Estimated Variances in the Determination of Individual Bioequivalence," Submitted.

Endrenyi, L., and L. Tothfalusi, 1999, "Subject-by-Formulation Interaction in Determination of Individual Bioequivalence: Bias and Prevalence," *Pharm, Res.*, 16:186-8.

Esinhart, J.D., and V.M. Chinchilli, 1994, "Sample Size Considerations for Assessing Individual Bioequivalence Based on the Method of Tolerance Interval," *Int. J. Clin. Pharmacol. Therap.*, 32(1):26-32.

Esinhart, J.D., and V.M. Chinchilli, 1994, "Extension to the Use of Tolerance Intervals for the Assessment of Individual Bioequivalence," *J. Biopharm. Stat.*, 4(1):39-52.

Fleiss, J.L., 1989, "A Critique of Recent Research on the Two-Treatment Crossover Design," *Controlled Clinical Trials*, 10:237-43.

Graybill, F., and C.M. Wang, 1980, "Confidence Intervals on Nonnegative Linear Combinations of Variances," *J. Amer. Stat. Assoc.*, 75:869-73.

Hauck, W.W., and S. Anderson, 1984, "A New Statistical Procedure for Testing Equivalence in Two-Group Comparative Bioavailability Trials," *J. Pharmacokin. Biopharm.*, 12:83-91.

Hauck, W.W., and S. Anderson, 1992, "Types of Bioequivalence and Related Statistical Considerations," *Int. J. Clin. Pharmacol. Therap.*, 30:181-7.

Hauck, W.W., and S. Anderson, 1994, "Measuring Switchability and Prescribability: When is Average Bioequivalence Sufficient?," *J. Pharmacokin. Biopharm.*, 22:551-64.

Hauck, W.W., M.-L. Chen, T. Hyslop, R. Patnaik, D. Schuirmann, and R.L. Williams for the FDA Population and Individual Bioequivalence Working Group, 1996, "Mean Difference vs. Variability Reduction: Tradeoffs in Aggregate Measures for Individual Bioequivalence," *Int. J. Clin. Pharmacol. Therap.*, 34:535-41.

Holder, D.J., and F. Hsuan, 1993, "Moment-Based Criteria for Determining Bioequivalence," *Biometrika*, 80:835-46.

Holder, D.J., and F. Hsuan, 1995, "A Moment-Based Method for Determining Individual Bioequivalence," *Drug Inf. J.*, 29:965-79.

Howe, W.G., 1974. "Approximate Confidence Limits on the Mean of X+Y Where X and Y are Two Tabled Independent Random Variables," *J. Amer. Stat. Assoc.*, 69:789-94.

Hsu, J.C., J.T.G. Hwang, H.-K. Liu, and S.J. Ruberg, 1994, "Confidence Intervals Associated with Tests for Bioequivalence," *Biometrika*, 81:103-14.

Hwang, S., P.B. Huber, M. Hesney, and K.C. Kwan, 1978, "Bioequivalence and Interchangeability,"*J. Pharm. Sci.*, 67:IV "Open Forum."

Hyslop, T., F. Hsuan, and D.J. Holder, "A Small-Sample Confidence Interval Approach to Assess

Individual Bioequivalence," Submitted.

Kimanani, E.K., and D. Potvin, 1998, "A Parametric Confidence Interval for a Moment-Based Scaled Criterion for Individual Bioequivalence," *J. Pharm. Biopharm.*, 25:595-614.

Liu, J.-P, 1995, "Use of the Repeated Crossover Designs in Assessing Bioequivalence," *Stat. Med.*, 14:1067-78.

Patnaik, R. N., L.J. Lesko, M.-L. Chen, R.L. Williams, and the FDA Population and Individual Bioequivalence Working Group, 1997, "Individual Bioequivalence: New Concepts in the Statistical Assessment of Bioequivalence Metrics," *Clin. Pharmacokin.*, 33:1-6.

Schall, R.A., 1995, "Unified View of Individual, Population and Average Bioequivalence," in *Bio-International 2: Bioavailability, Bioequivalence and Pharmacokinetic Studies* (H.H.Blume, K.K.Midha, eds.), Medpharm Scientific Publishers, 91-105.

Schall, R., and H.G. Luus, 1993, "On Population and Individual Bioequivalence," *Stat. Med.*, 12:1109-24.

Schall R., 1995, "Assessment of Individual and Population Bioequivalence Using the Probability That Bioavailabilities Are Similar," *Biometrics,* 51:615-26.

Schall, R., and R.L. Williams for the FDA Individual Bioequivalence Working Group, 1996, "Towards a Practical Strategy for Assessing Individual Bioequivalence," *J. Pharmacokin. Biopharm,* 24:133-49.

Schuirmann, D.J., 1987, "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Bioequivalence of Average Bioavailability," *J. Pharmacokin. Biopharm.*, 15:657-80.

Schuirmann, D.J., 1989, "Treatment of Bioequivalence Data: Log Transformation," in *Proceedings of Bio-International '89 - Issues in the Evaluation of Bioavailability Data*, Toronto, Canada, October 1-4, 159-61.

Senn, S., and D. Lambrou, 1998, "Robust and Realistic Approaches to Carry-Over," *Stat. Med.*, 17:2849-64.

Sheiner, L. B., 1992, "Bioequivalence Revisited," *Stat. Med.*, 11:1777-88.

Ting, N., R.K. Burdick, F.A. Graybill, S. Jeyaratnam, and T.F.C. Lu, 1990, "Confidence Intervals on

Linear Combinations of Variance Components That Are Unrestricted in Sign," *J. Stat. Comp. Sim.*, 35:135-43.

Westlake, W.J., 1973, "The Design and Analysis of Comparative Blood-Level Trials," in *Current Concepts in the Pharmaceutical Sciences, Dosage Form Design and Bioavailability* (J. Swarbrick, ed.), Lea and Febiger, 149-79.

Westlake, W.J., 1979, "Statistical Aspects of Comparative Bioavailability Trials," *Biometrics,* 35:273-80.

Westlake, W.J., 1981, "Response to Kirkwood, TBL.: Bioequivalence Testing - A Need to Rethink," *Biometrics,* 37:589-94.

Westlake, W.J., 1988, "Bioavailability and Bioequivalence of Pharmaceutical Formulations," in *Biopharmaceutical Statistics for Drug Development* (K.E. Peace, ed.), Marcel Dekker, Inc., 329-52.

APPENDIX A

Standards

The equations in section V call for standards to be established (i.e., σ_{T0} and θ_P for assessment of population BE, σ_{W0} and θ_I for individual BE). The recommended approach to establish these standards is described below.

A. σ_{T0} and σ_{W0}

As indicated in section IV, a general objective in assessing BE should be to compare the difference in the BA log-measure of interest after the administration of T and R formulation, T-R, with the difference in the same log-metric after two administrations of the R formulation, R-R′.

1. Population Bioequivalence

For population BE, the comparisons of interest should be expressed in terms of the ratio of the expected squared difference between T and R (administered to different individuals) and the expected squared difference between R and R′ (two administrations of R to different individuals), as shown below.

$$E(T - R)^2 = (\mu_T - \mu_R)^2 + \sigma_{TT}^2 + \sigma_{TR}^2 \qquad \textbf{Equation 8}$$

$$E(R - R')^2 = 2\sigma_{TR}^2 \qquad \textbf{Equation 9}$$

$$\frac{E(T - R)^2}{E(R - R')^2} = \frac{(\mu_T - \mu_R)^2 + \sigma_{TT}^2 + \sigma_{TR}^2}{2\sigma_{TR}^2} \qquad \textbf{Equation 10}$$

The population BE criterion in equation 4 is derived from equation 10, such that the criterion equals zero for two identical formulations.

The square root of equation 10 yields:

$$(\mu_T - \mu_R)^2 + \sigma_{TT}^2 + \sigma_{TR}^2$$

$$PDR = \left[\frac{}{2\sigma_{TR}^2}\right]^{1/2} \qquad \textbf{Equation 11}$$

The comparison between T and R expressed by equation 11 can be termed population difference ratio (PDR). The PDR is interpreted as the ratio of the expected T-R difference compared to the expected R-R' difference in the population. It should be noted that the PDR is monotonically related to the population BE criterion (PBC) described in equation 4 as follows:

$$PDR = (PBC/2 + 1)^{1/2} \qquad \textbf{Equation 12}$$

This guidance recommends that $\sigma_{T0} = 0.2$ based on the consideration of a maximum allowable PDR of 1.25.

2. Individual Bioequivalence

For individual BE, the comparisons of interest should be expressed in terms of the ratio of the expected squared difference between T and R (administered to the same individual) and the expected squared difference between R and R′ (two administrations of R to the same individual), as shown below.

$$E(T - R)^2 = (\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2 \qquad \textbf{Equation 13}$$

$$E(R - R')^2 = 2\sigma_{WR}^2 \qquad \textbf{Equation 14}$$

$$\frac{E(T - R)^2}{E(R - R')^2} = \frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2}{2\sigma_{WR}^2} \qquad \textbf{Equation 15}$$

The individual BE criterion in equation 6 is derived from equation 15, such that the criterion equals zero for two identical formulations. The square root of equation 15 is

$$IDR = \left[\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2}{2\sigma_{WR}^2}\right]^{1/2} \qquad \textbf{Equation 16}$$

The comparison of T and R expressed by equation 16 can be termed individual difference ratio (IDR). The IDR is interpreted as the ratio of the expected T-R

difference to the expected R-R' difference within an individual. Again, the IDR is monotonically related to the individual BE criterion (IBC) described in equation 6 as follows:

$$IDR = (IBC/2 + 1)^{1/2} \qquad \textbf{Equation 17}$$

This guidance recommends that $\sigma_{W0} = 0.2$, based on the consideration of the maximum allowable IDR of 1.25.

B. θ_P and θ_I

The determination of θ_P and θ_I should be based on the consideration of average BE criterion and the addition of variance terms to the population and individual BE criteria, as expressed by the formula below.

$$\theta = \frac{\text{average BE limit + variance factor}}{\text{variance}}$$

1. Population Bioequivalence

$$\theta_P = \frac{(\ln 1.25)^2 + \epsilon_P}{\sigma_{T0}^2} \qquad \textbf{Equation 18}$$

The value of ϵ_P for population BE is guided by the consideration of the variance term $(\sigma_{TT}^2 - \sigma_{TR}^2)$ added to the average BE criterion. This guidance recommends that $\epsilon_p = 0.02$.

2. Individual Bioequivalence

$$\theta_I = \frac{(\ln 1.25)^2 + \epsilon_I}{\sigma_{W0}^2} \qquad \textbf{Equation 19}$$

The value of ϵ_I for individual BE is guided by the consideration of the estimate of subject-by-formulation interaction (σ_D) as well as the difference in within-subject variability $(\sigma_{WT}^2 - \sigma_{WR}^2)$ added to the average BE criterion. As with the population BE

standard, the recommended allowance for the variance term ($\sigma_{WT}^2 - \sigma_{WR}^2$) is 0.02. In addition, this guidance recommends a σ_D^2 allowance of 0.03. The magnitude of σ_D is associated with the percentage of individuals whose average T to R ratios lie outside 0.8-1.25. It is estimated that if $\sigma_D = 0.1356$, ~10% of the individuals would have their average ratios outside 0.8-1.25, even if $\mu_T - \mu_R = 0$. When $\sigma_D = 0.1741$, the probability is ~20%.

Accordingly, on the basis of consideration for both σ_D and variability ($\sigma_{WT}^2 - \sigma_{WR}^2$) in the criterion, this guidance recommends that $\epsilon_I = 0.05$.

APPENDIX B

Choice of Specific Replicated Crossover Designs

Appendix B describes why we prefer replicated crossover designs with only two sequences, and why we recommend the specific designs described in section VI.A of this guidance.

1. Reasons Unrelated to Carryover Effects

Each unique combination of sequence and period in a replicated crossover design can be called a *cell* of the design. For example, the two-sequence, four-period design recommended in section VI.A has 8 cells. The four-sequence, four-period design below

		Period			
		1	2	3	4
Sequence	1	**T**	**R**	**R**	**T**
	2	**R**	**T**	**T**	**R**
	3	**T**	**T**	**R**	**R**
	4	**R**	**R**	**T**	**T**

has 16 cells. The total number of degrees-of-freedom attributable to comparisons among the cells is just the number of cells minus one (unless there are cells with no observations).

The fixed effects that are usually included in the statistical analysis are sequence, period, and treatment (i.e., formulation). The number of degrees-of-freedom attributable to each fixed effect is generally equal to the number of levels of the effect, minus one. Thus, in the case of the two-sequence, four-period design recommended in section V.A, there would be 2-1=1 degree-of-freedom due to sequence, 4-1=3 degrees-of-freedom due to period, and 2-1=1 degree-of-freedom due to treatment, for a total of 1+3+1=5 degrees-of-freedom due to the three fixed effects. Because these 5 degrees-of-freedom do not account for all 7 degrees-of-freedom attributable to the eight cells of the design, the fixed effects model is not *saturated.* There could be some controversy as to whether a fixed effects model that accounts for more or all of the degrees-of-freedom due to cells (i.e., a more saturated fixed effects model) should be used. For example, an effect for sequence-by-treatment interaction might be

included in addition to the three *main effects* - sequence, period, and treatment. Alternatively, an effect sequence-by-period interaction might be included, which would fully saturate the fixed effects model.

If the replicated crossover design has only two sequences, use of only the three main effects (sequence, period, and treatment) in the fixed effects model or use of a more saturated model makes little difference, provided there are no missing observations and the study is carried out in one group of subjects. The least squares estimate of μ_T-μ_R will be the same for the main effects model and for the saturated model. Also, the method of moments (MM) estimators of the variance terms in the model used in some approaches to assessment of population and individual BE (see Appendix G), which represent within-sequence comparisons, are generally fully efficient regardless of whether the main effects model or the saturated model is used.

If the replicated crossover design has more than two sequences, these advantages are no longer present. Main effects models will generally produce different estimates of μ_T-μ_R than saturated models (unless the number of subjects in each sequence is equal), and there is no well-accepted basis for choosing between these different estimates. Also, MM estimators of variance terms will be fully efficient only for saturated models, while for main effects models fully efficient estimators would have to include some between-sequence components, complicating the analysis. Thus, use of designs with only two sequences minimizes or avoids certain ambiguities due to the method of estimating variances or due to specific choices of fixed effects to be included in the statistical model.

2. Reasons Related to Carryover Effects

One of the reasons to use the four-sequence, four-period design described above is that it is thought to be optimal if carryover effects are included in the model. Similarly, the two-sequence, three-period design

		Period		
		1	2	3
Sequence	1	**T**	**R**	**R**
	2	**R**	**T**	**T**

is thought to be optimal among three-period replicated crossover designs. Both of these designs are *strongly balanced for carryover effects*, meaning that each treatment is preceded by each other treatment *and itself* an equal number of times.

With these designs, no efficiency is lost by including *simple* first-order carryover effects in the statistical model. However, if the possibility of carryover effects is to be considered in the statistical analysis of BE studies, the possibility of direct-by-carryover interaction should also be considered. If direct-by-carryover interaction is present in the statistical model, these favored designs are no longer optimal. Indeed, the TRR/RTT design does not permit an unbiased within-subject estimate of μ_T-μ_R in the presence of general direct-by-carryover interaction.

The issue of whether a purely main effects model or a more saturated model should be specified, as described in the previous section, also is affected by possible carryover effects. If carryover effects, including direct-by-carryover interaction, are included in the statistical model, these effects will be partially confounded with sequence-by-treatment interaction in four-sequence or six-sequence replicated crossover designs, but not in two-sequence designs.

In the case of the four-period and three-period designs recommended in section IV.A.1, the estimate of μ_T-μ_R, adjusted for first-order carryover effects including direct-by-carryover interaction, is as efficient or more efficient than for any other two-treatment replicated crossover designs.

3. Two-Period Replicated Crossover Designs

For the majority of drug products, two-period replicated crossover designs such as the Balaam design (which uses the sequences TR, RT, TT, and RR) should be avoided for individual BE because subjects in the TT or RR sequence do not provide any information on subject-by-formulation interaction. However, the Balaam design may be useful for particular drug products (e.g., a long half-life drug for which a two-period study would be feasible but a three- or more period study would not).

APPENDIX C

Sample Size Determination

Sample sizes for average BE should be obtained using published formulas. Sample sizes for population and individual BE should be based on simulated data. The simulations should be conducted using a default situation allowing the two formulations to vary as much as 5% in average BA with equal variances and certain magnitude of subject-by-formulation. The study should have 80 or 90% power to conclude BE between these two formulations. Sample size also depends on the magnitude of variability and the design of the study. Variance estimates to determine the number of subjects for a specific drug can be obtained from the biomedical literature and/or pilot studies.

The tables below give sample sizes for 80% and 90% power using the specified study design, given a selection of within-subject standard deviations (natural log scale), between-subject standard deviations (natural log scale), and subject-by-formulation interaction, as appropriate.

Average Bioequivalence
Estimated Recommended Numbers of Subjects
Δ=0.05

		80% Power		90% Power	
$\sigma_{WT} = \sigma_{WR}$	σ_D	2P	4P	2P	4P
0.15	0.01	12	6	16	8
	0.10	14	10	18	12
	0.15	16	12	22	16
0.23	0.01	24	12	32	16
	0.10	26	16	36	20
	0.15	30	18	38	24
0.30	0.01	40	20	54	28
	0.10	42	24	56	30
	0.15	44	26	60	34
0.50	0.01	108	54	144	72
	0.10	110	58	148	76
	0.15	112	60	150	80

Notes: 1. Results for two-period designs use method of Diletti et al. (Diletti 1991).
2. Four-period results use relative efficiency results of Liu (Liu 1995)

Population Bioequivalence
Four-Period Design (RTRT/TRTR)
Estimated Recommended Numbers of Subjects
ϵ_P =0.02, Δ=0.05

$\sigma_{WR} = \sigma_{WT}$	$\sigma_{BR}=\sigma_{BT}$	80% Power	90% Power
0.15	0.15	18	22
	0.30	24	32
0.23	0.23	22	28
	0.46	24	32
0.30	0.30	22	28
	0.60	26	34
0.50	0.50	22	28
	1.00	26	34

Note: Results for population BE are approximate from simulation studies (1,540 simulations for each parameter combination), assuming two-sequence, four-period trials with a balanced design across sequences.

Individual Bioequivalence
Estimated Recommended Numbers of Subjects
ϵ_I =0.05, Δ=0.05

		80% Power		90% Power	
$\sigma_{WT} = \sigma_{WR}$	σ_D	3P	4P	3P	4P
0.15	0.01	14	10	18	14
	0.10	18	14	24	18
	0.15	28	20	36	26
0.23	0.01	42	24	54	30
	0.10	56	30	74	40
	0.15	76	42	100	54
0.30	0.01	52	28	70	36
	0.10	60	32	82	42
	0.15	76	42	100	54
0.50	0.01	52	30	70	36
	0.10	60	32	82	42
	0.15	76	42	100	54

Note: Results for individual BE are approximate using simulations. The designs used in simulations are RTR/TRT (3P2S) and RTRT/TRTR (4P2S).

APPENDIX D

Rationale for Logarithmic Transformation of Pharmacokinetic Data

A. Clinical Rationale

The FDA Generic Drugs Advisory Committee recommended in 1991 that the primary comparison of interest in a BE study was the ratio, rather than the difference between average parameter data from T and R formulations. Using logarithmic transformation, the general linear statistical model employed in the analysis of BE data allows inferences about the difference between the two means on the log scale, which can then be retransformed into inferences about the ratio of the two averages (means or medians) on the original scale. Logarithmic transformation thus achieves a general comparison based on the ratio rather than the differences.

B. Pharmacokinetic Rationale

Westlake observed that a multiplicative model is postulated for pharmacokinetic measures in BA/BE studies (i.e., AUC and Cmax, but not Tmax) (Westlake 1973 and 1988). Assuming that elimination of the drug is first-order and only occurs from the central compartment, the following equation holds after an extravascular route of administration:

$$AUC_{0-\infty} = FD/CL \qquad \textbf{Equation 20}$$

$$= FD/(VK_e) \qquad \textbf{Equation 21}$$

where F is the fraction absorbed, D is the administered dose, and FD is the amount of drug absorbed. CL is the clearance of a given subject that is the product of the apparent volume of distribution (V) and the elimination rate constant (K_e).[4] The use of AUC as a measure of the amount of drug absorbed involves a multiplicative term (CL) that might be regarded as a function of the subject. For this reason, Westlake contends that the subject effect is not additive if the data is analyzed on the original scale of measurement.

---

[4] Note that a more general equation can be written for any multi-compartmental model as

$$AUC_{0-\infty} = FD/V_{d\beta}\,\lambda_n \qquad \textbf{Equation 22}$$

where $V_{d\beta}$ is the volume of distribution relating drug concentration in plasma or blood to the amount of drug in the body during the terminal exponential phase, and λ_n is the terminal slope of the concentration-time curve.

Logarithmic transformation of the AUC data will bring the CL (VK_e) term into the following equation in an additive fashion:

$$\ln AUC_{0-\infty} = \ln F + \ln D - \ln V - \ln K_e \qquad \textbf{Equation 23}$$

Similar arguments were given for Cmax. The following equation applies for a drug exhibiting one compartmental characteristics:

$$C_{max} = (FD/V) \times e^{-ke*Tmax} \qquad \textbf{Equation 24}$$

where again F, D and V are introduced into the model in a multiplicative manner. However, after logarithmic transformation, the equation becomes

$$\ln C_{max} = \ln F + \ln D - \ln V - K_e T_{max} \qquad \textbf{Equation 25}$$

Thus, log transformation of the Cmax data also results in the additive treatment of the V term.

APPENDIX E

SAS Program Statements for Average BE Analysis of Replicated Crossover Studies

The following illustrates an example of program statements to run the average BE analysis using PROC MIXED in SAS version 6.12, with SEQ, SUBJ, PER, and TRT identifying sequence, subject, period, and treatment variables, respectively, and Y denoting the response measure (e.g., log(AUC), log(Cmax)) being analyzed:

```
PROC MIXED;
CLASSES SEQ SUBJ PER TRT;
MODEL   Y = SEQ PER TRT/ DDFM=SATTERTH;
RANDOM  TRT/TYPE=FA0(2) SUB=SUBJ G;
REPEATED/GRP=TRT SUB=SUBJ;
ESTIMATE 'T vs. R' TRT 1 -1/CL ALPHA=0.1;
```

The Estimate statement assumes that the code for the T formulation precedes the code for the R formulation in sort order (this would be the case, for example, if T were coded as 1 and R were coded as 2). If the R code precedes the T code in sort order, the coefficients in the Estimate statement would be changed to -1 1.

In the Random statement, TYPE=FA0(2) could possibly be replaced by TYPE=CSH. This guidance recommends that TYPE=UN not be used, as it could result in an invalid (i.e., not non-negative definite) estimated covariance matrix.

Additions and modifications to these statements can be made if the study is carried out in more than one group of subjects.

APPENDIX F

Method for Statistical Test of Population Bioequivalence Criterion

Four-Period Crossover Designs

Appendix F describes a method for testing the population bioequivalence criterion (see section IV.B, equations 4 and 5). The procedure involves the computation of a test statistic that is either positive (does not conclude population bioequivalence) or negative (concludes population bioequivalence).

Consider the following statistical model that assumes a four-period design with equal replication of T and R in each of *s* sequences with an assumption of no (or equal) carryover effects (equal carryovers go into the period effects)

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl}$$

where i=1,..., s indicates sequence, j=1, ..., n_i indicates subject within sequence i, k= R, T indicates treatment, *l*=1, 2 indicates replicate on treatment k for subjects within sequence i. Y_{ijkl} is the response of replicate *l* on treatment k for subject j in sequence i, γ_{ikl} represents the fixed effect of replicate *l* on treatment k in sequence i, δ_{ijk} is the random subject effect for subject j in sequence i on treatment k, and ϵ_{ijkl} is the random error for subject j within sequence i on replicate *l* of treatment k. The ϵ_{ijkl}s are assumed to be mutually independent and identically distributed as

$$\epsilon_{ijkl} \sim N(0, \sigma_{wk}^2),$$

for i=1,...s, j=1, ..., n_i, k=R, T, and *l*=1, 2. Also, the random subject effects $\delta_{ij}=(\delta_{ijR}, \delta_{ijT})'$ are assumed to be mutually independent and distributed as

$$\delta_{ij} \sim N_2\left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BT}^2 \end{pmatrix}\right].$$

This model definition can be extended to other crossover designs.

Linearized Criteria:

$$\eta_1 = (\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2) - \theta_P \cdot \sigma_{TR}^2 < 0, \quad \text{for } \sigma_{TR} > \sigma_{T0}$$

$$\eta_2 = (\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2) - \theta_P \cdot \sigma_{T0}^2 < 0, \quad \text{for } \sigma_{TR} \le \sigma_{T0}$$

Estimating the Linearized Criteria:

The estimation of the linearized criteria depends on designs. The remaining estimation and confidence interval procedures assume a four-period design with equal replication of T and R in each of *s* sequences. The reparametrizations are defined as:

$$U_{Tij} = \tfrac{1}{2} * (Y_{ijT1} + Y_{ijT2})$$
$$U_{Rij} = \tfrac{1}{2} * (Y_{ijR1} + Y_{ijR2})$$
$$V_{Tij} = \tfrac{1}{\sqrt{2}} * (Y_{ijT1} - Y_{ijT2})$$
$$V_{Rij} = \tfrac{1}{\sqrt{2}} * (Y_{ijR1} - Y_{ijR2})$$

The variances of U_{Tij}, U_{Rij}, V_{Tij}, and V_{rij} are computed, pooling across sequences, and these variance estimates are denoted by MU_T, MU_R, MV_T, and MV_R, respectively. Then the linearized criteria are estimated by

$$\hat{\eta}_1 = \hat{\Delta}^2 + MU_T + 0.5 \cdot MV_T - (1+\theta_P) \cdot [MU_R + 0.5 \cdot MV_R],$$
$$\text{for } (MU_R + 0.5 \cdot MV_R)^{1/2} > \sigma_{T0}$$

$$\hat{\eta}_2 = \hat{\Delta}^2 + MU_T + 0.5 \cdot MV_T - (1+\theta_P) \cdot [MU_R + 0.5 \cdot MV_R] - \theta_P \cdot \sigma_{T0},$$
$$\text{for } (MU_R + 0.5 \cdot MV_R)^{1/2} \le \sigma_{T0}$$

To test for population bioequivalence, compute the 95% upper confidence bond of either the reference-scaled or constant-scaled linearized criterion. The procedure for computing this is described in the next paragraph. If this upper bound is negative, conclude population bioequivalence. If the upper bound is positive, do not conclude population bioequivalence.

95% Upper Confidence Bounds for Components:

Compute an upper confidence bound for $|\hat{\Delta}|$,

$$\left(\hat{\Delta}\right| + z_{1-\alpha,\beta} \cdot s\left(\frac{1}{s^2}\sum_{i=1}^{s} n_i^{-1} M_I\right)^{1/2}\Bigg)$$

For $(\hat{\Delta})^2 = E0$, H0= the square of the bound obtained for $|\hat{\Delta}|$, which is described above.

For $MU_1 = E1$, compute $H1 = \dfrac{(n-s)\cdot E1}{\chi^2_{n-s,\alpha}}$

For $0.5 \cdot MV_1 = E2$, compute $H2 = \dfrac{(n-s)\cdot E2}{\chi^2_{n-s,\alpha}}$

For $-(\hat{\theta}_R - \theta_R) MU_R$=E3rs, compute $H3rs = \dfrac{(n-s)\cdot E3rs}{\chi^2_{n-s,1-\alpha}}$

For $1 \cdot MU_R$=E3cs, compute $H3cs = \dfrac{(n-s)E3cs}{\chi^2_{n-s,1-\alpha}}$

For $(1+\theta_R)\cdot 0.5 \cdot MV_1 = E4rs$, compute $H4rs = \dfrac{(n-s)\cdot E4rs}{\chi^2_{n-s,1-\alpha}}$

For $0.5 \cdot MV_R$=E4cs, compute $H4cs = \dfrac{(n-s)E4cs}{\chi^2_{n-s,1-\alpha}}$

For each component above, also compute $Ui = (Hi - Ei)^2$. Note $n = \sum_{i=1}^{s} n_i$, where s is the number of sequences and n_i is the number of subjects per sequence.

95% Upper Confidence Bounds For Linearized Criteria:

$$H_{T1} = (E0 + E1 + E2 + E3rs + E4rs) - (U0 - U1 + U2 + U3rs + U4rs)^{1/2}$$

$$H_{T2} = (E0 - E1 - E2 + E3cs - E4cs - \theta_R \sigma_{T0}^3) + (U0 + U1 + U2 - U3cs + U4cs)^{1/2}$$

APPENDIX G

Variance Estimation

Relatively simple unbiased estimators, the method of moments (MM) or the restricted maximum likelihood (REML) method, can be used to estimate the mean and variance parameters in the individual BE criterion. A key distinction between the REML and MM methods relates to differences in estimating variance terms. The REML method estimates each of the three variances, σ_D^2, σ_{WR}^2, σ_{WT}^2, separately and then combines them in the individual BE criterion. The REML estimate of σ_D^2 is found from estimates of σ_{BR}^2, σ_{BT}^2, and the correlation, ρ. The MM approach is to estimate the sum of the variance terms in the numerator of the criterion, $\sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2$, and does not necessarily estimate each component separately. One consequence of this difference is that the MM estimator of σ_D^2 is unbiased but could be negative. The REML approach can also lead to negative estimates, but if the covariance matrix of the random effects is forced to be a proper covariance matrix, the estimate of σ_D^2 can be made to be non-negative. This forced non-negativity has the effect of making the estimate positively biased and introduces a small amount of conservatism to the confidence bound. The REML method can be used in special cases (e.g., when substantial missing data are present). In addition, the MM approaches have not yet been adapted to models that allow assessment of carryover effects.

APPENDIX H

Method for Statistical Test of Individual Bioequivalence Criterion

Appendix H describes a method for testing the individual bioequivalence criterion (see section IV.C, equations 6 and 7). The procedure involves the computation of a test statistic that is either positive (does not conclude individual bioequivalence) or negative (concludes individual bioequivalence).

Consider the following statistical model that assumes a four-period design with equal replication of T and R in each of *s* sequences with an assumption of no (or equal) carryover effects (equal carryovers go into the period effects)

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl}$$

where i=1,..., s indicates sequence, j=1, ..., n_i indicates subject within sequence i, k=R, T indicates treatment, *l*=1, 2 indicates replicate on treatment k for subjects within sequence i. Y_{ijkl} is the response of replicate *l* on treatment k for subject j in sequence i, γ_{ikl} represents the fixed effect of replicate *l* on treatment k in sequence i, δ_{ijk} is the random subject effect for subject j in sequence i on treatment k, and ϵ_{ijkl} is the random error for subject j within sequence i on replicate *l* of treatment k. The ϵ_{ijkl}s are assumed to be mutually independent and identically distributed as

$$\epsilon_{ijkl} \sim N(0, \sigma_{Wk}^2),$$

for i=1,...s, j=1, ..., n_i, k=R, T, and *l*=1, 2. Also, the random subject effects $\delta_{ij}=(\delta_{ijR}, \delta_{ijT})'$ are assumed to be mutually independent and distributed as

$$\delta_{ij} \sim N_2\left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BT}^2 \end{pmatrix}\right].$$

This model definition can be extended to other crossover designs.

Linearized Criteria:

$$\eta_1 = (\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2) - \theta_I \cdot \sigma_{WR}^2 < 0, \quad \text{for } \sigma_{WR} > \sigma_{W0}$$

$$\eta_2 = (\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2) - \theta_I \cdot \sigma_{W0}^2 < 0, \quad \text{for } \sigma_{WR} \le \sigma_{W0}$$

Estimating the Linearized Criteria:

The estimation of the linearized criteria depends on designs. The remaining estimation and confidence interval procedures assume a four-period design with equal replication of T and R in each of *s* sequences. Define the reparametrizations:

$$I_{ij} = Y_{ijT1} - Y_{ijR1}$$

$$T_{ij} = Y_{ijT1} - Y_{ijT2}$$

$$R_{ij} = Y_{ijR1} - Y_{ijR2}$$

Compute the variances of I_{ij}, T_{ij}, and R_{ij}, pooling across sequences, and denote these variance estimates by M_I, M_T, and M_R, respectively. Then the linearized criteria are estimated by

$$\hat{\eta}_1 = \hat{\Delta}^2 + M_I + 0.5\,M_T - (1.5 + \theta_I)\,M_R, \quad \text{for } (M_R)^{1/2} > \sigma_{W0}$$

$$\hat{\eta}_2 = \hat{\Delta}^2 + M_I + 0.5\,M_T - 1.5\,M_R - \theta_I\,\sigma_{W0}^2, \quad \text{for } (M_R)^{1/2} \le \sigma_{W0}$$

To test for individual bioequivalence, compute the 95% upper confidence bound of either the reference-scaled or constant-scaled linearized criterion. The procedure for computing this is described in the next paragraph. If this upper bound is negative, conclude individual bioequivalence. If the upper bound is positive, do not conclude individual bioequivalence.

95% Upper Confidence Bounds for Criteria:

The table below illustrates the construction for the two-sequence, four-period design for the reference-scaled criterion, $\hat{\eta}_1$:

H_q=Confidence Limit	E_q=Point Estimate	$U_q=(H_q - E_q)^2$
$H_D = \left(\left\| \hat{\Delta} \right\| + t_{1-\alpha,n-s} \left(\frac{1}{s^2} \sum_{i=1}^{s} n_i^{-1} M_I \right)^{1/2} \right)^2$	$E_D = \hat{\Delta}^2$	U_D
$H_I = \frac{((n-s) \cdot M_I)}{\chi^2_{\alpha,n-s}}$	$E_I = M_I$	U_I
$H_T = \frac{(0.5\ (n-s)\ M_T)}{\chi^2_{\alpha,n-s}}$	$E_T = 0.5 \cdot M_T$	U_T
$H_R = \frac{(-(1.5+\theta_I)\ (n-s)\ M_R)}{\chi^2_{1-\alpha,n-s}}$	$E_R = -(1.5+\theta_I) \cdot M_R$	U_R
$H_{\eta_1} = \sum E_q + \left(\sum U_q \right)^{1/2}$		

where $n - \sum_{i=1}^{s} n_i$, and s is the number of sequences. $H_{\eta_1} = \sum E_q + \left(\sum U_q \right)^{1/2}$ is the upper 95% confidence bound for $\hat{\eta}_1$. The confidence limit for $\hat{\eta}_2$ is computed similarly, adjusting the constants associated with the variance components where appropriate.